

## Automation of Qualitative Content Analysis: A Proposal

*Annette Hoxtell*

### Key words:

qualitative content analysis;  
automation; semi-automation;  
digitization;  
artificial intelligence;  
categories; coding frame; text mining; transcription

**Abstract:** (Semi-)automation of qualitative content analysis will likely occur in the foreseeable future; in fact, it is already partially possible. Based upon my own qualitative content analysis, I provide an overview of the forms semi-automated qualitative content analysis might take, that includes the additional steps of data collection and transcription. This analysis remains a semi-automated one, since it depends on continuous human-machine interaction. I call attention to automation approaches that are already in use and seem suited to supporting progress towards the (semi-)automation of qualitative content analysis.

### Table of Contents

- [1. Introduction](#)
- [2. Initial Qualitative Content Analysis—As Carried Out](#)
- [3. Qualitative Content Analysis and Automation](#)
- [4. \(Semi-\)Automation of Qualitative Content Analysis](#)
  - [4.1 Preliminary work: Transcription](#)
  - [4.2 Primary work: Qualitative content analysis](#)
  - [4.3 Extension: Data collection and communicative validation](#)
- [5. Conclusion and Outlook](#)

### [References](#)

### [Author](#)

### [Citation](#)

## 1. Introduction

From 2014 to 2015, I interviewed 72 people as part of my dissertation research, transcribed the interviews, and evaluated them using qualitative content analysis. For months on end, I was lost in the depths of analysis: Did the coding framework represent all relevant passages? Did the division into key categories and subcategories make sense? What was I supposed to do with passages that fit into more than one (sub)category? I would have loved to have been able to shorten this repetitive and time-consuming process; from time to time I would have even loved to hand it over to somebody—or should I say something?—else. I thought it should be possible to automate qualitative content analysis. The automation idea stuck with me in the years that followed, during which I carried out further qualitative content analyses, and taught my students this method. [1]

In the next section, I will introduce my dissertation's research design, as well as the qualitative content analysis I originally conducted (Section 2), followed by the definitions of key terms (Section 3). In section four, I describe the forms semi-automated qualitative content analysis might take and outline the approaches to automation already in use. Finally, I discuss who might profit from semi-

automation of qualitative content analysis and the circumstances under which this might occur (Section 5). [2]

## 2. Initial Qualitative Content Analysis—As Carried Out

My dissertation design was explorative and made pragmatic use of the methods most suitable to my inquiry (STAMANN, JANSSEN & SCHREIER, 2016). My main research question was: For which reasons do youths choose a training company? Narrative interviews represented my core method of data collection since they served well to highlight youth experiences, opinions, and motives (CRESWELL, 2007; SCHÜTZE, 1976; WATTANASUWAN, BUBER & MEYER, 2007). Some of these interviews were a part of four multiple case studies (YIN, 2009), which also included expert interviews (BORTZ & DÖRING, 2016; PFADENHAUER, 2009), participant observation, and advertising media in order to triangulate data and methods of analysis (FLICK, 2011). [3]

I transcribed all the interviews verbatim (MAYRING, 2016) and conducted structured qualitative content analysis (KUCKARTZ, 2016) by adapting MAYRING's (2015) summarizing and structuring-deductive category assignment as well as STEIGLEDER's (2007) continuous revision of categories. Throughout the analysis, which adhered to strict rules and allowed for content to be consolidated and placed in thematical order (NADERER, 2007; SCHREIER, 2014), I continuously referred to the main research question. Figure 1 demonstrates my approach.

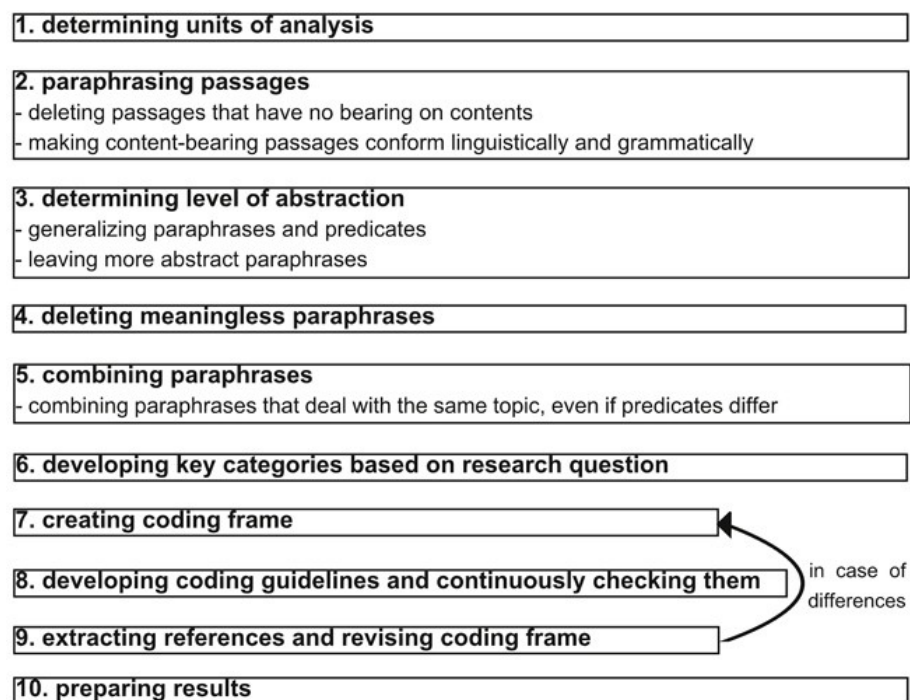


Figure 1: Process of summarizing-structuring content analysis (HOXTELL, 2016, p.42; inspired by MAYRING, 2010, pp.68-70; STEIGLEDER, 2007, p.183) [4]

### 3. Qualitative Content Analysis and Automation

Before describing existing approaches to the automation of qualitative content analysis, I will define the terms *qualitative content analysis*, *digitization*, and *automation*. *Qualitative content analysis* is a rule-based, interpretive method used to systematically generate categories from data (KUCKARTZ, 2016; NADERER, 2007; SCHREIER, 2014; STAMANN et al., 2016; STEIGLEDER, 2007). A significant number of researchers develop and use qualitative content analysis methods, which differ in name and the ways in which categories are developed (BOYATZIS, 1998; HSIEH & SHANNON, 2005; KUCKARTZ, 2016; MAYRING, 2000, 2015; SCHREIER, 2012, 2014). Usually qualitative content analysis is applied to written data. It differs from other types of content analysis, such as dictionary-based, computer-assisted content analysis in two major ways. Firstly, analysts consider context while interpreting data, i.e., how the data came into existence and which effects it has (MAYRING, 2015). Secondly, ambiguities are permitted (KUCKARTZ, 2016). [5]

*Digitization* is a prerequisite for the automation of qualitative content analysis. Digitization means, on the one hand, that analogue content is transformed into digital content and, on the other hand, that machines process digital content (WEBER & BUSCHBACHER, 2017). Content analysis often takes place in a digitally assisted form, e.g., when computer programs such as [ATLAS.ti](#), [audiotranskription](#), or [MAXQDA](#) independently count the frequency of particular categories in various data sets. [6]

Using such programs does not amount to full *automation* of qualitative content analysis. The term *automation* signifies that computational systems take over tasks originally carried out by humans, often controlling processes and making decisions (VOIGT, n.d.). Sensors, for example, control the flow of oil or gas in pipelines using ultrasound and report leaks or blockages (GEIPEL-KERN, 2011), before the human eye can perceive them. A fully automated qualitative content analysis would have to generate new categories on its own, when the data requires it to do so.<sup>1</sup> [7]

Fully automated qualitative content analysis would constitute a type of *weak artificial intelligence* since it would adapt to changing conditions, instead of just following rules (NATIONAL SCIENCE AND TECHNOLOGY COUNCIL, COMMITTEE ON TECHNOLOGY, 2016; PEINL, 2018a; SCHMIECH, 2018). [8]

---

<sup>1</sup> Automation takes over human tasks. Therefore, I use anthropomorphisms deliberately.

## 4. (Semi-)Automation of Qualitative Content Analysis

### 4.1 Preliminary work: Transcription

Transcription, often the first step in qualitative content analysis, annoyed me. Why did I have to type out 72 interviews? And if I did not do the typing, why should I have to pay somebody else to do it? Why was there no automatic transcription? Other social scientists seem to share my sentiment, as the questions regularly popping up on mailing list QSF-L suggested. On 23 November 2017, a researcher wrote, "Has anyone of you used direct language recognition for interviews? Or is subsequent transcription still the silver bullet?"<sup>2</sup> [9]

I want a system that detects the type of data, no matter whether they are in audio, video, or text form (GÖBEL, 2018). It should identify the speaker according to sound patterns (BITKOM, 2018) and transcribe what was said. Currently, computational systems are good at detecting different types of data, as well as patterns within them. This applies, for example, to spam filters, graphological software and self-driving cars (CLARK, 2016; FEHRENBACHER, 2015; LÜTTERS, 2017; PAKALSKI, 2006). Voice-operated assistance systems, such as Alexa, Bixby, Cortana, Google Home, Siri, and Watson seem especially suited to transcription purposes. They can differentiate human voices from ambient noise, to react to requests (LÜTTERS, 2018), and to learn from them. [10]

Dictation software such as [Dragon](#) or the [Dictate](#) feature in Microsoft Office are well suited for transcription, though only for a single speaker. Since 2019, [f4transkript](#) has transformed conversations in German between two or more people into text. If there are only two speakers, manual correction of the automatically generated transcript takes less time than manual transcription. [11]

### 4.2 Primary work: Qualitative content analysis

I started my qualitative content analysis by reading parts of the interviews, noting down what came to mind, and underlining passages that seemed relevant to code (KUCKARTZ, 2016). Thus, I *determined the units of analysis* one by one. Even in this phase, I would have loved to have had a computational system execute these steps. This would have allowed me to compare notes and underlined passages. Subsequently, I could have *determined the level of abstraction*, and I would have informed the computational system of the units of analysis with which I agreed and the ones with which I did not agree. Next, I would have *paraphrased* some *passages* and the computational intelligence system would have paraphrased the rest. It would have also *combined paraphrases* and *deleted meaningless ones*. The computational system would have documented the proceedings in a way similar to the manually developed ones in Table 1.

---

2 Subscribers to QSF-L can access the mail at [https://lists.fu-berlin.de/private/qsf\\_l/2017-November/msg00041.html](https://lists.fu-berlin.de/private/qsf_l/2017-November/msg00041.html). Registration is open at [https://lists.fu-berlin.de/listinfo/qsf\\_l](https://lists.fu-berlin.de/listinfo/qsf_l).

Case, Line	Passage	Paraphrase	Combination
1, 12	Trustworthy people who accept me for who I am	To be viewed as a whole human being at work instead of a resource	K 4 Staff Relationship Human image
3, 25	I believe I will be friends with my colleagues. I cannot stand to pretend that we don't know each other, even though we work together. Understanding makes life easier.	Friendly relationship	
7, 29	They were interested in me. They wanted to know something about me. What I do in my spare time. Why I chose this occupation.	To be viewed and accepted as a whole human being at work	

Table 1: Paraphrasing and combination [12]

This would have been the moment for me to evaluate the computational system's work: I would have accepted or rejected paraphrases and their combinations, which would have been similar to working with the review feature in Microsoft Word. The computational system would have made its coding decisions transparent (BUNDESREGIERUNG, 2018) by documenting them (BITKOM, 2018), as well as the reasons behind those decisions. [13]

I *developed key categories* and *created a coding frame* based upon my research question. The section displayed in Table 1 refers to the key category "training company" and its subcategory "staff." I *developed coding guidelines* illustrating the coding frame. Table 2 shows a portion of the coding guidelines.

Category	Definition	Anchor Example	Coding Rules
1.3.1 Reasons for the Company	Reasons for the training company phrased in broad terms	I want to feel comfortable. Everything is connected. Everything must feel right.	Reasons that do not belong in a specific subcategory
1.3.1.1 Company Climate	Working environment, social interaction at work	Social, not too stuffy, attentive	A description of the company's characteristics and what they do for the student
1.3.1.2 Salary	Salary	Hm, after all money is important.	How much the interviewees want to earn
1.3.1.3 Fun	Fun and/or joy at work	I want to enjoy work.	Intrinsic motivation of apprentices: enjoying work and interactions there

Table 2: Coding guideline [14]

Once again, I would have loved for a machine to execute these steps for the first interviews. I could have compared coding guidelines and accepted or rejected categories and their relationships to one another. [15]

Based on my review of the machine's work, it could have *extracted references* from the second batch of interviews and *continuously revised* key categories and coding guidelines. Where ambiguities arose, it could have paused its analysis and asked me to make the relevant decisions (GÖBEL, 2018; WEBER & BUSCHBACHER, 2017). I would have had the ability to enter any part of the coding process and alter or undo it. [16]

Currently, full automation of qualitative content analysis does not exist. There are however automated processes that have the capability to assess meaning and develop key categories. Units of analysis may be determined through *text mining*, a technique used to explore text in bibliographic databases such as MEDLINE (PAPANIKOLAOU, PAVLOPOULOS, THEODOSIOU & ILIOPOULOS, 2015; WESTERGAARD, STÆRFELDT, TØNSBERG, JENSEN & BRUNAK, 2018) and the German National Library (JUNGER & SCHWENS, 2017). Text mining is also used to determine genres in online bookselling (WEBER & BUSCHBACHER, 2017), which can be considered key categories in the terminology of qualitative content analysis. [17]

Social scientists have used automated frequency and difference analysis to develop key categories from online posts on the German federal government's civil dialogue webpage. They, then, validated those categories manually and added subcategories (WALDHERR et al., 2019). Manual validation amounts to human control, a prerequisite for many types of applications of artificial intelligence. Human-machine interaction may also function the other way around. Doctors, for example, sometimes seek an automated second opinion in cancer prevention practices. Together, humans and machines achieve a better diagnosis (WANG, KHOSLA, GARGEYA, IRSHAD & BECK, 2016). Judicata, a legal search engine, asks for human advice in case of uncertainty. It also reveals its reasoning (GURARI, 2017). *Sentiment analysis* gives an overview of the tones of posts; it is, for example, used to differentiate between positive, negative, and neutral customer reviews (BANNISTER, 2018; KREUTZER & SIRRENBURG, 2019). Textifier qualitative data analysis software forms categories semi-automatically; however, it depends upon human groundwork and validation (PRIEBE, 2016; R., 2019; please see, e.g., [DiscoverText](#)). [18]

#### 4.3 Extension: Data collection and communicative validation

In addition to transcription and qualitative content analysis, I would like to automate certain types of data collection. Instead of talking to each young person individually, I could have automated later interviews. At the beginning of each interview, participants could have answered open-ended questions which could have immediately been analyzed by the computational system. Later on, participants would have been confronted with closed-ended questions based upon responses provided by previous participants. As a result, data collection would have yielded new responses, and this would have allowed for the quantification of all responses. The analysis would not need to be text-based; a computational system could also analyze speech and issue results in audio format. [19]

Voice-operated assistance systems enable automated speech input and output in market research. Participants can listen to recordings, read transcripts, and adjust them (LÜTTERS, 2017). The system demands communicative validation (FLICK, 2002)—in an auditory or visual form—and, thus, increases its internal validity. [20]

Table 3 gives an overview of the steps proposed above to semi-automate qualitative content analysis. The process could be adjusted to accommodate other types of qualitative content analysis.

Step	Human	Machine
Preliminary work: Transcription	- controls transcripts and adjusts them as needed	- detects type of data - recognizes who is speaking - transcribes
Defining units of analysis	- for all texts - compares units of analysis as defined by human and machine and adjusts them as needed	- for all texts
Paraphrasing + determining level of abstraction + combining paraphrases + deleting paraphrases	- performs this for sections of texts - controls and adjusts	- does so for all texts
Developing key categories + creating coding frame + developing coding guidelines	- performs this based on a portion of all texts - intervenes in case of ambiguity - reviews coding and makes adjustments as needed	- does so for all texts - continuously revises coding guidelines and framework - ceases analysis in cases of ambiguity and requests human input - documents procedure
Processing results	- does so completely	
Extension: Data collection and communicative validation	- performs first interviews - controls and adjusts as needed	- performs further interviews - performs qualitative content analysis in real time - poses closed-ended questions based on results from previous interviews - performs quantitative analysis - requests communicative validation

Table 3: Outline of a semi-automated qualitative content analysis [21]



## 5. Conclusion and Outlook

People who conduct qualitative content analysis individually and are not a part of a larger research project with multiple coders, may benefit from semi-automation of qualitative content analysis. A computational system can stimulate researchers to reflect upon the ways in which they code. This occurs because the system codes all data, unlike human participants in research workshops who only examine certain sections of data. [22]

Automation of qualitative content analysis may also be beneficial for large research projects that depend on large quantities of data being analyzed according to uniform rules. The reason for this is that computational systems need to process a lot of data before they can learn and adapt (PEINL, 2018a). Independent of a research project's size, all researchers would benefit from an *open data* strategy when automating qualitative content analysis. Large quantities of high-quality data from the sciences and public services could be made available from decentralized sources and in compliance with the law, as demanded by, for example, the German government (BUNDESREGIERUNG, 2018). Data could be transferred for specific purposes through open interfaces (WEBER & BUSCHBACHER, 2017). This would make it harder for data collecting and processing entities to concentrate power and acquire monopolies (PEINL, 2018b). [23]

The coding process and coding guidelines could be disclosed (SCHARKOW, 2012), as could subsequently coded passages. The latter should however be made public anonymously. Plagiarism detection software and Fraunhofer-Gesellschaft's Medical Data Space already disclose sensitive data in a way that conforms with data privacy protections. When plagiarism detection software, such as [PlagScan](#), detects passages that have been copied or paraphrased from unpublished works, but not referenced, the software displays relevant passages from original works in anonymized form. It is nonetheless possible to contact the person who uploaded the original work. Fraunhofer-Gesellschaft's Medical Data Space is based on data administered in a decentralized manner, that is only exchanged as needed (WEBER & BUSCHBACHER, 2017; please refer to [Fraunhofer Medical Data Space](#), too). In 2016, the [Rotterdam Exchange Format Initiative](#) was founded to set an interoperability standard for the exchange of data. Thanks to this initiative, qualitative data can be processed using various qualitative data analysis programs. This includes qualitative content analysis, as well. [24]

My opinion is that humans should interact with machines to benefit from the semi-automation of qualitative content analysis. Both can play on their strengths: Humans are better than machines at hermeneutic interpretation; machines are better than humans at processing large amounts of data. Portions of this kind of semi-automated analysis already exist; they only need to be combined and refined to comprise a complete semi-automated system, for which there are two requirements. Firstly, machines must account for their operations in a way that is comprehensible to humans. Secondly, humans must be able to intervene in the

processes in which machines engage. Due to this division of labor and the potential need for human intervention, qualitative content analysis may only be semi-automated. Full automation is not feasible now; without human intervention, it would not be qualitative. [25]

## References

- Bannister, Kristian (2018). Understanding sentiment analysis: What it is & why it's used. *Blog post*, February 26, <https://www.brandwatch.com/blog/understanding-sentiment-analysis/> [Date of Access: June 21, 2019].
- BITKOM (2018). *Digitalisierung gestalten mit dem Periodensystem der Künstlichen Intelligenz*. Berlin: Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V., [https://bitkom.de/sites/default/files/2018-12/181204\\_LF\\_Periodensystem\\_online\\_0.pdf](https://bitkom.de/sites/default/files/2018-12/181204_LF_Periodensystem_online_0.pdf) [Date of Access: January 21, 2019].
- Bortz, Jürgen & Döring, Nicola (2016). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Boyatzis, Richard E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, CA: Sage.
- Clark, Andrew (2016). Machine Learning mit der Crowd bietet viele Vorteile. *Blog post*, October 30, <https://www.crowdguru.de/blog/machine-learning-mittels-crowd/> [Date of Access: April 20, 2018].
- Creswell, John W. (2007). *Qualitative inquiry & research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage.
- Bundesregierung (2018). *Eckpunkte der Bundesregierung für eine Strategie Künstliche Intelligenz*, [https://www.bmbf.de/files/180718%20Eckpunkte\\_KI-Strategie%20final%20Layout.pdf](https://www.bmbf.de/files/180718%20Eckpunkte_KI-Strategie%20final%20Layout.pdf) [Date of Access: January 22, 2019].
- Fehrenbacher, Katie (2015). How Tesla is ushering in the age of the learning car. *Fortune*, October 16, <http://fortune.com/2015/10/16/how-tesla-autopilot-learns/> [Date of Access: April 20, 2018].
- Flick, Uwe (2002). *Qualitative Sozialforschung. Eine Einführung*. Hamburg: Rowohlt.
- Flick, Uwe (2011). *Triangulation: eine Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Geipel-Kern, Anke (2011). *Bilder: Zehn Beispiele aus der Praxis – Wie Automatisierung die Energieeffizienz verbessert*, <https://www.process.vogel.de/index.cfm?pid=7513&pk=333136&fk=0&type=article#2> [Date of Access: January 29, 2019].
- Göbel, Richard (2018). Big Data – Herausforderung bei der Nutzung großer Datenmengen. In Dietmar Wolff & Richard Göbel (Eds.), *Digitalisierung: Segen oder Fluch? Wie die Digitalisierung unsere Lebens- und Arbeitswelt verändert* (pp.73-99). Berlin: Springer.
- Gurari, Itai (2017). Embedding humans in the loop. *Judicata Blog post*, November 17, <https://blog.judicata.com/embedding-humans-in-the-loop-2c19fb41c926> [Date of Access: January 28, 2019].
- Hoxtell, Annette (2016). Entscheidungsgründe für den Ausbildungsbetrieb – eine explorative Studie mit Auszubildenden sowie Schülerinnen und Schülern. *Dissertation*, Technical and economic education, Potsdam University, Germany, <http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-93377> [Date of Access: June 13, 2019].
- Hsieh, Hsiu-Fang & Shannon, Sarah E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288.
- Junger, Ulrike & Schwens, Ute (2017). Die inhaltliche Erschließung des schriftlichen kulturellen Erbes auf dem Weg in die Zukunft. *DNB professionell*, [https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Erschliessen/inhaltlicheErschliessungZukunft.pdf?\\_\\_blob=publicationFile&v=2](https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Erschliessen/inhaltlicheErschliessungZukunft.pdf?__blob=publicationFile&v=2) [Date of Access: June 12, 2019].
- Kreutzer, Ralf & Sirrenberg, Marie (2019). *Künstliche Intelligenz verstehen: Grundlagen – Use-Cases – unternehmenseigene KI-Journey*. Wiesbaden: Springer.
- Kuckartz, Udo (2016). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (3rd ed.). Weinheim: Beltz Juventa.
- Lütters, Holger (2017). Sprich mit mir!. *planung & analyse*, 17(5), 58-60.

Lütters, Holger (2018). Digitale Sprachassistenten: in der Forschung. Von der Technologie zur Anwendung. In Matthias Knaut (Ed.), *KREATIVITÄT + X = INNOVATION Beiträge und Positionen der HTW Berlin* (pp.34-41). Berlin: BWV.

Mayring, Philipp (2000). Qualitative content analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(2), Art. 20, <https://doi.org/10.17169/fqs-1.2.1089> [Date of Access: August 4, 2019].

Mayring, Philipp (2010). *Qualitative Inhaltsanalyse Grundlagen und Techniken* (11th ed.). Weinheim: Beltz.

Mayring, Philipp (2015). *Qualitative Inhaltsanalyse Grundlagen und Techniken* (12th ed.). Weinheim: Beltz.

Mayring, Philipp (2016). *Einführung in die qualitative Sozialforschung: eine Anleitung zu qualitativem Denken* (6th ed.). Weinheim: Beltz.

Naderer, Gabriele (2007). Auswertung & Analyse qualitativer Daten. In Gabriele Naderer & Eva Balzer (Eds.), *Qualitative Marktforschung in Theorie und Praxis: Grundlagen, Methoden und Anwendungen* (pp.363-391). Wiesbaden: Gabler.

National Science and Technology Council, Committee on Technology (2016). *Preparing for the future of artificial intelligence*. Washington, D.C.: Executive Office of the President, <https://info.publicintelligence.net/WhiteHouse-ComputationalIntelligencePreparations.pdf> [Date of Access: January 23, 2019].

Pakalski, Ingo (2006). Strato: Filter erkennt Bilder-Spam. *golem.de*, December 14, <https://www.golem.de/0612/49471.html> [Date of Access: April 20, 2018].

Papanikolaou, Nikolas; Pavlopoulos, Georgios A.; Theodosiou, Theodos & Iliopoulos, Ioannis (2015). Protein-protein interaction predictions using text mining methods. *Methods*, 74, 47-53.

Peinl, René (2018a). Privatleben 4.0 – Wie Digitalisierung, das Internet der Dinge und Deep Learning unser Privatleben verändern. In Dietmar Wolff & Richard Göbel (Eds.), *Digitalisierung: Segen oder Fluch? Wie die Digitalisierung unsere Lebens- und Arbeitswelt verändert* (pp.225-252). Berlin: Springer.

Peinl, René (2018b). Smart Speaker ohne Cloud. *Video*, December 14, <https://www.youtube.com/watch?v=3UXzxghpcuo> [Date of Access: January 23, 2019].

Pfadenhauer, Michaela (2009). Das Experteninterview: Ein Gespräch auf gleicher Augenhöhe. In Renate Buber & Hartmut H. Holzmüller (Eds.), *Qualitative Marktforschung: Konzepte – Methoden – Analysen* (pp.449-461). Wiesbaden: Gabler.

Priebe, John (2016). Best practices for classifying uncoded data. *Texifter*, April 4, <https://texifter.zendesk.com/hc/en-us/articles/217965828-Best-practices-for-classifying-uncoded-data> [Date of Access: June 21, 2019].

R., Iris (2019). Classify a large volume of data quickly. *Texifter*, February 6, <https://texifter.zendesk.com/hc/en-us/articles/202972680-Classify-a-large-volume-of-data-quickly> [Date of Access: June 21, 2019].

Scharkow, Michael (2012). *Automatische Inhaltsanalyse und maschinelles Lernen*. Berlin: epubli.

Schmiech, Chris (2018). Der Weg zur Industrie 4.0 für den Mittelstand – Ausgewählte Potenziale und Herausforderungen. In Dietmar Wolff & Richard Göbel (Eds.), *Digitalisierung: Segen oder Fluch? Wie die Digitalisierung unsere Lebens- und Arbeitswelt verändert* (pp.1-28). Berlin: Springer.

Schreier, Margrit (2012). *Qualitative content analysis in practice*. London: Sage.

Schreier, Margrit (2014). Ways of doing qualitative content analysis: Disentangling terms and terminologies. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 15(1), Art. 18, <http://dx.doi.org/10.17169/fqs-15.1.2043> [Date of Access: February 1, 2016].

Schütze, Fritz (1976). Zur Hervorlockung und Analyse von Erzählungen thematisch relevanter Geschichten im Rahmen soziologischer Feldforschung: dargestellt an einem Projekt zur Erforschung von kommunalen Machtstrukturen. In Ansgar Weymann & Arbeitsgruppe Bielefelder Soziologen (Eds.), *Kommunikative Sozialforschung: Alltagswissen und Alltagshandeln, Gemeindemachtforschung* (pp.159-260). München: Fink.

Stamann, Christoph; Janssen, Markus & Schreier, Margrit (2016). Searching for the core: Defining qualitative content analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 17(3), Art. 16, <http://dx.doi.org/10.17169/fqs-17.3.2581> [Date of Access: April 20, 2018].

Steigleder, Sandra (2007). *Die strukturierende qualitative Inhaltsanalyse im Praxistest: eine konstruktiv kritische Studie zur Auswertungsmethodik von Philipp Mayring*. Marburg: Tectum.

Voigt, Kai-Ingo (o.D.). Definition: Automatisierung. *Gabler Wirtschaftslexikon*, <https://wirtschaftslexikon.gabler.de/definition/automatisierung-27138/version-250801> [Date of Access: January 22, 2019].

Waldherr, Annie; Wehden, Lars-Ole; Stoltenberg, Daniel; Miltner, Peter; Ostner, Sophia & Pfetsch, Barbara (2019). Inductive codebook development for content analysis: Combining automated and manual Methods. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 20(1), Art. 19, <http://dx.doi.org/10.17169/fqs-20.1.3058> [Date of Access: January 29, 2019].

Wang, Dayong; Khosla, Aditya; Gargeya, Rishab; Irshad, Humayun & Beck, Andrew H. (2016). Deep learning for identifying metastatic breast cancer. *arXiv:1606.05718* [cs, q-bio], <https://arxiv.org/pdf/1606.05718.pdf> [Date of Access: January 28, 2019].

Wattanasuwan, Kritsadarat; Buber, Renate & Meyer, Michael (2007). Das narrative Interview und die narrative Analyse. In Uwe Flick (Ed.), *Qualitative Sozialforschung: eine Einführung* (pp.359-379). Hamburg: Rowohlt.

Weber, Mathias & Buschbacher, Florian (2017). *Künstliche Intelligenz – Wirtschaftliche Bedeutung, gesellschaftliche Herausforderungen, menschliche Verantwortung*. Berlin: BITKOM, DFKI, <https://www.bitkom.org/sites/default/files/file/import/171012-KI-Gipfelpapier-online.pdf> [Date of Access: January 22, 2019].

Westergaard, David; Stærfeldt, Hans-Henrik; Tønsgaard, Christian; Jensen, Lars Juhl & Brunak, Søren (2018). A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts. *PLOS Computational Biology*, 14(2), <https://doi.org/10.1371/journal.pcbi.1005962> [Date of Access: June 13, 2019].

Yin, Robert K. (2009). *Case study research: Design and methods* (4th ed.). Los Angeles, CA: Sage.

## Author

Annette HOXTELL is professor of business studies at hwtk – University of Applied Sciences Berlin.

She does qualitative research in marketing and related areas and teaches marketing and intercultural communication. Annette HOXTELL has worked in the IT industry as a PR manager and sustainability advisor. She also ran a mentoring project to increase the number of girls involved in STEM disciplines.

Contact:

Prof. Dr. Annette Hoxtell

hwtk Berlin  
Bernburger Straße 24/25, 10965 Berlin,  
Germany

Tel.: +49 30 206176-74

Fax: +49 30 206176-71

E-Mail: [annette.hoxtell@hwtk.de](mailto:annette.hoxtell@hwtk.de)

URL: <https://www.hwtk.de/prof-dr-annette-hoxtell/>

## Citation

Hoxtell, Annette (2020). Automation of Qualitative Content Analysis: A Proposal [25 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 20(3), Art. 15, <http://dx.doi.org/10.17169/fqs-20.3.3340>.